

Архивиране и компресиране на файлове в Линукс

ас. Стоян Мечев

катедра „Информационни технологии“

ВВМУ „Н. Й. Вапцаров“

Архивиране и компресиране

- Архивирането е процес на групиране на много файлове в един.
- Компресирането е пренаписването на данните по такъв начин, че резултатният файл заема по-малко място в сравнение с оригиналния.
 - Компресия без загуба на данни - възстановява се напълно идентично копие на оригинала, извлечено от компресирания обект.
 - ABVBBAACSSSSCAAAABA AAAAC
 - A*4BAA*5C*4AB*5AC (run-length encoding)
 - Компресия със загуба на данни JPEG, MP3.
- Варианти
 - обединяване на много файлове в един .tar файл и след това компресиране
 - '+' алгоритъмът за компресиране може да мине през няколко файла и така се постига по-ефективна компресия;
 - '-' при повреден архив, всички данни от точката на повредата са неизползваеми;
 - нищо не ви спира да си компресирате много файлове по отделно и да ги обедините в .tar файл.

Архивиране на файлове с командата tar

- Името на командата идва от "tape archive"
 - Записва файловете един след друг;
 - Добавя информация – дата, собственик, права за достъп и др.;
 - Може да бъде записван върху различни носители;
 - Стандартен формат за разпространение на изходния код за Linux и други свободни софтуерни пакети.
 - Архивите обикновено не са компресирани
 - GNU tar има по-разширени възможности в сравнение с tar за други Unix варианти
- Командите **split** и **cat**
 - **split -d -b 10k pacman-1.3.tar.gz**
 - **cat x00 x01 x02 x03**
- Предимства: надежден, има го на всички Unix и Linux системи.
- Недостатъци: не всички версии поддържат файлове за устройства (/dev/sda1 и д.р.)
- По подразбиране tar архивите не са компресирани.
 - .tar, tar.bz2, tar.gz

Командата tar - опции

Традиционно tar {A|c|d|r|t|u|x}[GnSkUW0mpsMBiajJzZhPlRvwo] [ARG...]

- Опции
 - -c ("create") създава нов архив
 - -f file създава новия архив на (или чете съществуващ архив от) файл, където "файл" може да бъде обикновен файл или файл с устройство (наред с други)
 - -M обработва многотомни архиви
 - -r добавя файлове към архива (не за магнитна лента)
 - -t показва "съдържанието" на архива
 - -u замества файловете, които са по-нови от тяхната версия в архива. Ако файлът изобщо не е в архива, той се вмъква (не за магнитна лента)
 - -v (Verbose mode) - показва какво прави tar в момента
 - -x извлича файлове и директории от архив
 - -z компресира или декомпресира архива с помощта на gzip
 - -Z компресира или декомпресира архива с помощта на команда "compress" (обикновено не е достъпна за Linux)
 - -j компресира или декомпресира архива с помощта на bzip2
- Опции, които не приемат параметри могат да се групират в произволен ред.
- Опция, която приема параметър се добавя в края на групата.
- Можете да пропуснете тирето пред първата група опции, но не се препоръчва.
 - tar cvf

Командата tar - примери

- **tar -cvf ~/data.tar data*** - архивира всички файлове, чиито имена започват с "data" в архива data.tar, който се намира в домашната директория на потребителя.
 - Автоматично включва цялото съдържание на поддиректориите, ако има такива.
 - Преобразува абсолютните пътища в относителни.
- **tar -tvf data.tar** - извежда подробна информация за съдържанието на архива.
- **tar -xvf data.tar** - разархивира данните
 - може да се изваждат само отделни файлове от архива, но имената им трябва да съвпадат точно.
 - **tar -xvf ../test.tar README** - ще разархивира само файла README

Компресиране на файлове с gzip

- Използва се за компресиране на единични файлове. Добавя ".gz" към имената.
- Премества файл в архива.
- Разработена е за да се избегнат лицензионни претенции относно програмата **compress**, която ползва Lempel-Ziv-Welch (LZW) алгоритъм.
 - Само компресирането е обект на патент. Декомпресирането не е.
- **gzip** ползва Deflate метод разработен от Phil Katz.
- Да не се бърка с Windows програмите, които имат "ZIP" в името, като PKZIP, които едновременно компресират и архивират.

Опции за gzip

- -r (“recursive”) компресира файлове в поддиректории, пак всеки файл по отделно;
- -S <suffix> използва посочения суфикс вместо .gz
- -c (gzip -c ⇔ zcat) извежда компресирания файл на stdout. Оригиналът остава непроменен
- -d декомпресира файла (алтернативно: gunzip работи като gzip -d)
- -l (“list”) извежда справочна информация за компресирания файл, като например името на файла, оригинала и опакования размер;
- -v извежда името и коефициента на компресия на всеки файл, който вече е компресиран
- -1... -9 определя коефициент на компресия: -1 (или --fast) работи най-бързо, но не се компресира толкова добре, докато -9 (или --best) води до най-добрата компресия с по-бавна скорост; настройката по подразбиране е -6.

Примери

- **gzip my_file.txt** – компресира файла.
- **gzip my_file.txt.gz** – декомпресира файла ⇔ **gunzip my_file.txt.gz**
- **gzip -vS .t letter.tex** – компресира файла и заменя суфикса с ".t"
 - S опцията трябва да се използва и при декомпресирането, ако файлът е компресиран с нея.
- Използване заедно с **tar**
 - **tar -cvzf text-all.tar.gz *.txt** – добавя всички файлове с разширение ".txt" към архив text-all.tar и след това го компресира.
 - оригиналните файлове остават.
 - **tar -xvzf text-all.tar.gz** – декомпресира файла и след това го разархивира.

Компресиране на файлове с bzip2

- Съвместима с **tar** – използва се опцията **j**
- Използва друг метод (Burrows-Wheeler трансформация и след това Huffman схема)
- Опциите **-s** и **-d** са същите.
- **-1** до **-9** са други – определят размера на блока по време на компресията. По подразбиране е **-9** (отговаря на 900 K)
 - Изборът на размер на блок за компресиране оказва влияние върху паметта необходима за декомпресиране – декомпресирането на компютри с малко RAM може да се окаже проблем.

Архивиране и компресиране на файлове с използване на zip и unzip

- Използва се за изпращане на файлове до компютри под Windows
- Ако извикате **zip** със съществуващ ZIP архив като негов първи параметър и не посочите нищо друго, файловете за компресиране се добавят към компресирания архив върху съществуващото му съдържание (съществуващи файлове със същите имена се презаписват).
- **zip test.zip file1 file2** - добавя file1 и file2 към архива test.zip
- Опции
 - -r рекурсивно обхожда поддиректориите.
 - -u "обновява" компресирания архив чрез добавяне на файлове, само ако файлът, споменат в командния ред, е по-нов от съществуващ файл със същото име в компресирания архив (файловете, които все още не са в архива се добавят във всеки случай).
 - -f "освежава" компресирания архив - файловете се презаписват с по-нови версии от командния ред, но само ако те действително вече съществуват в архива (в архива не се добавят напълно нови файлове).
 - -d разглежда имената на файловете в командния ред като имена на файловете в компресирания архив и ги изтрива.

Примери за използване на unzip

- **unzip -v test** - справка за съдържанието на архива test.zip (разширението може да се пропусне);
- **unzip test** - разархивира файла test;
- **unzip test ziptest/file1** - извлича само file1 (директорията ще се създаде при необходимост)
- **unzip test -x ziptest/file1** - изключва file1 от разархивирането
- **unzip test "ziptest/f*"** - позволява използването на шел-шаблони за търсене на файлове.
 - обърнете вниманието на "" - за да може unzip да ползва шаблона;
 - за разлика от шаблоните на шел, се ползват пълните имена на файловете - включително "/".

- Благодаря за вниманието!
- Въпроси?